

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP05/050264

International filing date: 21 January 2005 (21.01.2005)

Document type: Certified copy of priority document

Document details: Country/Office: DE
Number: 10 2004 009 617.1
Filing date: 27 February 2004 (27.02.2004)

Date of receipt at the International Bureau: 24 March 2005 (24.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

BUNDESREPUBLIK DEUTSCHLAND

EP05/50264

**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

Aktenzeichen: 10 2004 009 617.1

Anmeldetag: 27. Februar 2004.

Anmelder/Inhaber: Siemens Aktiengesellschaft, 80333 München/DE

Bezeichnung: Verfahren und Vorrichtung zur Codierung und
Decodierung von strukturierten Dokumenten

IPC: G 06 K 17/21

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 7. März 2005
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag


Schärer

Beschreibung

Verfahren und Vorrichtung zur Codierung und Decodierung von strukturierten Dokumenten

5

Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur Codierung und Decodierung von strukturierten Dokumenten gemäß dem Oberbegriff des Anspruchs 1.

10 XML (= extensible markup language) ist eine Sprache, mit der eine strukturierte Beschreibung der Inhalte eines Dokuments mittels XML-Schema-Sprachdefinitionen ermöglicht wird. Eine genauere Beschreibung des XML-Schemas sowie der darin verwendeten Strukturen, Datentypen und Inhaltsmodelle findet sich
15 in den Referenzen

- <http://www.w3.org/TR/2001/REC-xmlschema-0-20010502/>,
- <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/>,
- <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/>.

20 Aus Schriften zum MPEG-7-Standard, insbesondere ISO/IEC 15938-1 "Multimedia Content Description Interface -Part 1: Systems", Geneva 2002, sind Verfahren, Vorrichtungen oder Systeme zur Codierung bzw. Decodierung von XML-basierten Dokument bekannt.

Erweiterungen zu dem Verfahren, Vorrichtungen oder Systeme zur Codierung bzw. Decodierung von XML-basierten Dokument sind aus Schriften zum MPEG-7-Standard, sind aus der deutschen Anmeldung mit dem amtlichen Aktenzeichen 10351897.5
30 bekannt. Darin ist ein Verfahren zur Codierung eines strukturierten Dokuments, insbesondere eines XML-basierten Dokuments offenbart, bei dem eine Vielzahl von Codes mittels eines oder mehrerer Schemas und/oder Namensräume erzeugt werden, wobei für ein Schema und/oder einen Namensraum und/oder für eine
35 Gruppe von Schemas und/oder Namensräumen jeweils separate, von anderen Schemas und/oder Namensräumen unabhängige Codes für die mittels in den Schemas und/oder Namensräumen und/oder

in den Gruppen von Schemas und/oder Namensräumen definierten und/oder deklarierten Elemente vergeben werden.

5 Diese ermöglichen eine effiziente Codierung auch dann, wenn Schemas dem Encoder und/oder Decoder nicht vollständig bekannt sind. Dies wird erreicht, indem Codetabellen für die Datentypen, die globalen Elemente und die Ersetzungsgruppen nach Namensräumen getrennt werden, wobei unter einem Namensraum dabei ein Raum zu verstehen ist, in dem darin verwendeten Namen von Datentypen (Typnamen) mit eindeutigen Bedeutungen belegt und definiert sind.

15 Bekannte Verfahren zur binären Repräsentation von MPEG-7 und anderen XML-basierten Beschreibungen oder Dokumenten weisen Defizite hinsichtlich En- und Decodierungskomplexität auf, sofern die zu codierende XML Beschreibung oder das XML Dokument auf mehreren Namensräumen beruht. Beispielsweise wird in den oben genannten Schriften ein Verfahren zur binären Repräsentation von XML-Beschreibungen und XML-Dokumenten beschrieben, das Codetabellen für XML-Beschreibungen und XML-Dokumente basierend auf Schemas und Namensräumen bestimmt (Im Folgenden wird der Ausdruck "Namensraum" synonym für den Ausdruck "Schema" verwendet).

Gemäß der bekannten Verfahrensweise können dabei Datentypen von anderen Datentypen vererbt sein. Diese Vererbungsbeziehung erlaubt, in einem XML Dokument anstatt einer Instanz des Basistyps eine Instanz eines vererbten Typs zu verwenden.

30 Der Typecode signalisiert ausgehend vom Basistyp welchen Typs eine Instanz ist. Ausgehend von einem Basistypen in einem ersten Namensraum muss zur Bestimmung der adressierbaren Typnamen in einem zweiten Namensraum bei En- und/oder Decodierung die Vererbungsstruktur über mehrerer Namensräume analysiert werden. Hierzu wird ein Vererbungsbaum aufgebaut, wie

35 er in der ISO/IEC 15938-1 "Multimedia Content Description Interface -Part 1: Systems", Geneva 2002, beschrieben ist.

- Dies setzt voraus, dass alle Namensräume bekannt sind und der gesamte Vererbungsbaum im Speicher aufgebaut werden kann. Der gesamte Vererbungsbaum besteht aus den Kennzeichnern der be-
- 5 nannten Typen aller für die Instanziierung einer XML Beschreibung und/oder XML Dokuments referenzierten Namensräume und deren Vererbungsbeziehung. Das beschriebene Verfahren ist daher sehr aufwendig.
- 10 Die der Erfindung zugrundeliegende Aufgabe besteht nun darin, ein gegenüber dem aus dem Stand der Technik vereinfachtes Verfahren und eine Vorrichtung zur Codierung und Decodierung von strukturierten Dokumenten anzugeben.
- 15 Diese Aufgabe wird ausgehend von dem Verfahren zur Codierung gemäß den Merkmalen des Oberbegriffs des Anspruchs 1, ausgehend von dem Verfahren zur Codierung gemäß den Merkmalen des Oberbegriffs des Anspruchs 2, ausgehend von dem Verfahren zum Decodieren gemäß den Merkmalen des Oberbegriffs des Anspruchs
- 20 17, sowie ausgehend von der Codiervorrichtung gemäß Anspruch 22, der Decodiervorrichtung gemäß Anspruch 23 und der Codier-/Decodiervorrichtung gemäß dem Anspruch 24, jeweils durch deren kennzeichnenden Merkmale gelöst.
- Bei dem erfindungsgemäßen Verfahren zur Codierung eines strukturierten, insbesondere XML-basierten Dokuments, bei dem eine Vielzahl von Codes mittels eines oder mehrerer Namens-
- 30 räume erzeugt und für mittels Namensräumen definierten Typen vergeben werden, wird eine Teilmenge adressierbarer Typen eines der Namensräume auf Grundlage von Vererbungsbeziehungen zwischen den Namensräumen sowie der Namensräume der Basistypen der Teilmenge ermittelt.
- 35 Das Verfahren zeichnet sich dadurch vorteilhaft aus, dass lediglich ein geringer Teil der insgesamt vorhandenen Namensräume zur Identifizierung der adressierbaren Teilmenge gespeichert

bzw. geladen werden muss. Daher ist eine enorme Entlastung von Ressourcen die Folge und zudem wird das Codieren beschleunigt.

- 5 Alternativ oder ergänzend wird bei dem Verfahren zur Codierung eines strukturierten, insbesondere XML-basierten, Dokuments zu jedem Namensraum eine Zuordnung zu weiteren Namensräumen derart gebildet wird, dass zumindest eine Zuordnungsinformation dergestalt erzeugt, dass zumindest eine Vererbungsbeziehung zwischen einem erbenden Namensraum und vererbenden Namensräumen beschrieben ist. Hierbei wird ein Namensraum, der Typen enthält, die direkt von einem Basistypen aus einem anderen Namensraum vererbt sind, erbender Namensraum genannt und ein Namensraum, der Basistypen enthält, die in
 10 einen anderen Namensraum vererbt worden sind, vererbender Namensraum.
 15

Die durch diese Weiterbildung bereitgestellten Zuordnungsinformationen ermöglichen eine strukturierte Organisation von
 20 Vererbungsinformationen, so dass lediglich ein Teil des gesamten Vererbungsbaums für die Identifizierung der Teilmenge von Nöten ist. Diese Weiterbildung führt somit zu einer weiteren Ressourceneinsparung/-entlastung und Beschleunigung.

Vorzugsweise wird die Zuordnungsinformation des erbenden Namensraums aus einer Liste von Codes der Basistypen von Kopf-
 typen des erbenden Namensraums, gebildet, wobei Kopftypen Typen sind, die direkt von einem Basistyp des vererbenden Namensraums abstammen und wobei die Basistypen auch durch Kopf-
 30 typen gebildet werden, denen weitere Kopftypen entstammen.

Bevorzugt wird die adressierbare Teilmenge ausgehend von einem Startbasistyp der Basistypen des vererbenden Namensraums ermittelt, wobei zur Identifizierung der Teilmenge in der Regel
 35 ausgehend von dem Startbasistyp zur Ermittlung der Teilmenge Kopftypen im erbenden Namensraum durch die Zuordnungsinformation identifiziert werden, die von einem Basistypen

aus dem vererbenden Namensraum abstammen, bei denen der Startbasistyp ein Basistyp in dem vererbenden Namensraum ist.

Alternativ oder ergänzend wird bei dem Verfahren zur Codierung eines strukturierten, insbesondere XML-basierten, Dokuments zu mindestens einem Namensraum die den ererbenden Namensräumen zugeordnete Zuordnungsinformation gemeinsam mit dem jeweiligen Namensraum in einem die Codierung und/oder Decodierung vornehmenden ersten Gerät gespeichert wird.

Bei einer Weiterbildung wird die den ererbenden Namensräumen zugeordnete Zuordnungsinformation in einem zweiten Gerät erzeugt und gemeinsam dem jeweiligen Namensraum in einem die Codierung und/oder Decodierung vornehmenden ersten Gerät übermittelt.

Das hier beschriebene Verfahren ist vorteilhaft, da nun zur Bestimmung der adressierbaren Datentypen nur der Namensraum des Basistypen, der Namensraum des zu adressierenden Datentyps und die Vererbungsbeziehung BT bekannt sein und/oder geladen werden muss.

Ein wesentlicher Vorteil des erfindungsgemäßen Verfahrens ist eine effiziente Bestimmung der adressierbaren Typnamen zu ermöglichen, ohne den gesamten Vererbungsbaum aufbauen zu müssen. Zudem kann dies auch ohne Kenntnis aller Schemas oder Namensräume erfolgen.

Ein weiterer Vorteil besteht darin, dass die Suche nach dem adressierten Datentyp mit weniger Vergleichoperationen realisiert werden kann verglichen mit der Suche im gesamten Vererbungsbaum.

In einer Ausführungsform besteht die Vererbungsinformation BT eines Namensraums NS aus einer Liste an Typcodes TC^{LBT} der Basistypen LBT je Kopftypen HT des Namensraums NS.

In einer weiteren Ausführungsform hierzu werden die Typecodes nach folgendem Verfahren vergeben:

5 Zur Codierung eines strukturierten Dokuments werden eine Vielzahl von Codes mittels eines oder mehrerer Schemas und/oder Namensräume erzeugt. Dabei werden für ein Schema und/oder einen Namensraum und/oder für eine Gruppe von Schemas und/oder Namensräumen jeweils separate, von anderen Schemas und/oder Namensräumen unabhängige Codes für die mittels
10 in den Schemas und/oder Namensräumen und/oder in den Gruppen von Schemas und/oder Namensräumen definierten und/oder deklarierten Elemente vergeben.

15 Bei Weiterbildung hiervon werden Codes in Schemas und/oder Namensräumen separiert vergeben. Das hier beschriebene Verfahren ist vorteilhaft, da nun Schemas und/oder Namensräume nach Bedarf auch während der Übertragung von Dokumenten geladen werden können und existierende Codetabellen für andere Namensräume sich hierdurch nicht ändern und somit nicht neu
20 erstellt werden müssen. Ein weiterer Vorteil besteht darin, dass die separaten Codes für Fälle, in denen sehr viele Namensräume importiert werden, weniger Bits zur Adressierung benötigen, als wenn wie in ISO/IEC 15938-1 "Multimedia Content Description Interface -Part 1: Systems", Geneva 2002, alle Namensräume zusammengefasst werden. Auch in Fällen, in denen ein sehr großer Namensraum importiert wird, können die separaten Codes für die anderen Namensräume mit weniger Bits codiert werden.

30 In einer bevorzugten Variante der Erfindung sind die separaten Codes in Adressbereiche eingeteilt, wobei über die Adressbereiche das Schema und/oder der Namensraum bzw. die Gruppe von Schemas und/oder Namensräumen identifizierbar wird.

35 In einer bevorzugten Ausführungsform des erfindungsgemäßen Codierverfahrens umfassen die separaten Codes jeweils einen

lokalen Code bezüglich des Schemas und/oder des Namensraums und/oder bezüglich der Gruppe von Schemas und/oder Namensräumen und einen Identifikationscode, der das Schema und/oder den Namensraum und/oder die Gruppe von Schemas und/oder Namensräumen identifiziert. Ein lokaler Code ist hierbei ein Code, der eindeutig innerhalb des durch den Identifikationscode identifizierten Schemas bzw. Namensraums ist.

Vorzugsweise werden separate Codes für globale Elemente und/oder SubstitutionGroups und/oder Datentypen vergeben. Eine genaue Definition für globale Elemente, SubstitutionGroups und Datentypen findet sich in den XML-Schema-Definitionen, die in den Dokumenten - <http://www.w3.org/TR/2001/REC-xmlschema-0-20010502/> , <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/> und <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/>, genau erläutert sind.

Für Datentypen TypeCodes, die in dem Dokument ISO/IEC 15938-1 "Multimedia Content Description Interface -Part 1: Systems", Geneva 2002, erläutert sind, werden separate Codes in einer bevorzugten Ausführungsform derart erzeugt, dass innerhalb des Vererbungsbaums eines Namensraums der zu einem ersten Datentyp in demselben Namensraum benachbarten Datentyp einen Codeabstand zu dem ersten Datentyp hat, welcher der Anzahl der in diesem Namensraum von dem ersten Datentyp abgeleiteten Datentypen entspricht. Ein Datentyp ist zu einem ersten Datentyp benachbart, wenn der Datentyp vom gleichen Basisdatentyp wie der erste Datentyp abgeleitet worden ist und dem Datentyp unter allen Datentypen, die von diesem Basisdatentypen abgeleitet worden sind, der kleinste TypeCode zugewiesen wurde, der größer als der TypeCode des ersten Datentyps ist. Bei dieser Ausführungsform werden die Codes für die Datentypen TypeCodes innerhalb des - möglicherweise disjunkten - Vererbungsbaums so vergeben, dass eine vorteilhafte Nachbarschaftsbeziehung in einem gegebenen Namensraum entsteht und erhalten bleibt, auch wenn in diesem Namensraum Unterbäume mit aus anderen Namensräumen abgeleiteten Typen vorkommen.

In einer besonders bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens werden die separaten Codes innerhalb eines gegebenen Namensraums gemäß einem Verfahren vergeben,
 5 das folgende Schritte umfasst:

- in einem ersten Schritt werden alle Datentypen eines Namensraums, die von Datentypen anderer Namensräume vererbt worden sind, in der im MPEG-7 Standard definierten Reihenfolge der globalen TypeCodes der jeweiligen Basisdatentypen in einer Liste sortiert, wobei die Basisdatentypen die Datentypen in anderen Namensräumen sind, von denen die sortierten Datentypen vererbt worden sind;
- 10 - in einem zweiten Schritt werden jeweils diejenigen Datentypen eines Namensraums, die von einem bestimmten Basisdatentypen eines bestimmten anderen Namensraums vererbt worden sind, lexikographisch sortiert;
- 15 - in einem dritten Schritt werden alle Datentypen eines Namensraums, die nicht von einem Datentypen eines anderen Namensraums vererbt worden sind, entsprechend der im MPEG-7 Standard definierten Reihenfolge in die bestehende Liste von Datentypen einsortiert;
- 20 - in einem vierten Schritt werden die separaten Codes in der Reihenfolge der Liste an die Datentypen des Namensraums vergeben.

Der Vorteil dieser Ausführungsform ist, dass der adressierte Datentyp, insbesondere ein TypeCode, schnell gefunden und somit decodiert werden kann. Nach den Regeln in ISO/IEC 15938-1
 30 "Multimedia Content Description Interface -Part 1: Systems", Geneva 2002, adressiert ein TypeCode einen abgeleiteten Typ relativ zu einem Basistypen. Also definiert der Basistyp einen Unterbaum, in dem alle adressierbaren Datentypen vorhanden sind. Sind in dem Unterbaum nun mehrere Namensräume enthalten, so kann durch die vorteilhafte Nachbarschaftsbeziehung, die durch die obige Ausführungsform der Erfindung erreicht wird, in dem Namensraum ein adressierter Datentyp
 35

schnell gefunden werden, da durch einen Vergleich eines gesuchten Datentyps mit zwei benachbarten Datentypen im sortierten Vererbungsbaum festgestellt werden kann, ob sich der gesuchte Datentyp im Unterbaum des Datentypen der zwei benachbarten Datentypen mit dem kleineren binären Code befindet. Auf diese Weise kann der Suchaufwand erheblich verringert werden. Ein weiterer Vorteil dieser Nachbarschaftsbeziehung besteht darin, dass ein Decoder bei einer Codierung der TypeCodes gemäß ISO/IEC 15938-1 "Multimedia Content Description Interface -Part 1: Systems", Geneva 2002, die Codewortlänge, die sich aus der Anzahl der abgeleiteten Datentypen bestimmt, direkt aus dem Codeabstand der benachbarten Datentypen berechnen kann.

In einer weiteren Ausführungsform werden die lokalen Typcodes nach dem oben beschriebenen Verfahren vergeben, wobei in einer Weiterbildung hiervon, der Typecode TC^{LBT} aus NamensraumID und lokalem Typcode nach dem in oben beschriebenen Verfahren gebildet wird.

In einer weiteren Ausführungsform werden die lokalen Typcodes nach dem oben beschriebenen Verfahren vergeben und nur Basistypen des ersten Namensraums betrachtet, deren lokaler Typcode

- a) größer als der lokale Typcode des Startbasistyps OBT ist und
- b) kleiner als der kleinste, nächstgrößere lokale Typcode eines zum Startbasistypen OBT benachbarten Typen ist.

In einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens werden die Vererbungsbeziehungen BT zwischen Namensräumen mit einem Schema und/oder Namensraum gespeichert und/oder übertragen.

Neben dem oben beschriebenen erfindungsgemäßen Codierverfahren betrifft die Erfindung ferner ein Decodierverfahren, mit dem ein strukturiertes Dokument, insbesondere ein XML-

basiertes Dokument decodiert wird, wobei das Verfahren derart ausgestaltet ist, dass ein mit dem erfindungsgemäßen Codierverfahren codiertes Dokument decodiert wird.

- 5 In einer bevorzugten Ausgestaltung des erfindungsgemäßen Decodierverfahrens wird hierbei zur Decodierung eines binären TypeCodes - dessen Erzeugung oben beschrieben ist - die Code-
länge des separaten Codes für den binären TypeCode aus der
Anzahl der abgeleiteten Daten bestimmt. Vorzugsweise wird
10 ferner in einer bevorzugten Ausführungsform zur Decodierung eines bestimmten TypeCodes des Subbaums des Vererbungsbaums des Namensraums, in dem sich der bestimmte TypeCode befindet, anhand der Codeabstände zwischen benachbarten Datentypen ermittelt.
- 15 Bei einer Weiterbildung wird zur Bestimmung der Basistypen die von einem Startbasistypen entstammen, anhand der Codeabstände zwischen benachbarten Datentypen ermittelt.
- 20 Eine weitere Alternative oder Ergänzung ist gegeben, wenn bei dem zur Bestimmung der Anzahl an Typen in der Teilmenge ausgehend von den Kopftypen anhand der Codeabstände zwischen benachbarten Kopftypen ermittelt wird.
- 30 Neben den oben beschriebenen Verfahren betrifft die Erfindung ferner eine Codiervorrichtung sowie eine Decodiervorrichtung zur Durchführung der erfindungsgemäßen Codier- bzw. Decodierverfahrens. Ferner umfasst die Erfindung eine Codier- und eine Decodiervorrichtung, mit der das erfindungsgemäße Codierverfahren und das erfindungsgemäße Decodierverfahren durchführbar ist.

Ausführungsbeispiele der Erfindung werden nachfolgend anhand der beigefügten Zeichnungen näher erläutert.

35

Es zeigen:

- Figur 1 eine Prinzipdarstellung eines Codier- und Decodier-systems in welchem das erfindungsgemäße Verfahren zum Tragen kommt.
- 5 Figur 2 eine Darstellung einer beispielhaften XML-Schema-Definition, in der auch Datentypen aus anderen Namensräumen importiert sowie abgeleitet werden.
- 10 Figur 3 eine Darstellung eines Vererbungsbaums von Datentypen, einschließlich der Zuordnung der lokalen Codes zu in den Namensräumen auftretenden Typen.
- 15 Figur 4 eine Darstellung eines Vererbungsbaums von Datentypen, der sich über mehrerer Namensräume erstreckt.
- Figur 5 eine Darstellung eines Vererbungsbaums inklusive Vererbungsinformationen zwischen Namensräumen.
- 20 In Figur 1 ist beispielhaft ein Codier- und Decodiersystem, in dem das erfindungsgemäße Verfahren zum Einsatz kommt, mit einem Encoder ENC und einem Decoder DEC dargestellt, mit denen XML-Dokumente DOC codiert bzw. decodiert werden. Sowohl der Encoder als auch der Decoder verfügen beide über ein sogenanntes XML-Schema S, in dem die zur Kommunikation genutzten Elemente und Typen des XML-Dokuments deklariert und definiert sind. Aus dem Schema S werden über entsprechende Schema-Compilationen SC im Encoder und Decoder Code-Tabellen CT erzeugt. Wenn das XML-Dokument DOC codiert wird, werden den
- 30 Inhalten des XML-Dokuments über die Code-Tabellen binäre Codes zugeordnet. Hierdurch wird eine Binärdarstellung BDOC des Dokuments DOC erzeugt, die mithilfe der Code-Tabelle CT im Decoder wieder decodiert werden kann. Es können hierbei mehrere Schemas verwendet werden, insbesondere können auch Schemas eingesetzt werden, welche auf einem Basisschema beruhen
- 35 und aus einem weiteren Schema abgeleitet werden.

In Figur 2 ist beispielhaft ein Auszug aus einer XML-Schema-Definition gezeigt. Dem Fachmann sind solche XML-Schema-Definitionen bekannt, so dass auf den genauen Inhalt des Auszugs der Figur 2 nicht eingegangen wird. Der Auszug enthält zwei Schemadefinitionen, zum einen wird im oberen Teil ein Schema A definiert, wie durch eine geschweifte Klammer angedeutet ist, zum anderen wird im unteren Teil ein Schema X definiert, wie ebenfalls durch eine geschweifte Klammer angedeutet wird. Das Schema X verwendet wiederum Datentypen, die aus dem Schema A importiert worden sind.

In Figur 3 sind die Vererbungsbeziehungen zwischen einem ersten Namensraum NS1 und einem zweiten Namensraum NS2 und deren Datentypen in der Form eines Ausschnitts einer Baumstruktur grafisch dargestellt. Wie der Figur anhand des rückwärts gerichteten nichtgestrichelten Pfeils entnommen werden kann, besteht eine Vererbungsbeziehung zwischen dem zweiten Namensraum NS2 und dem ersten Namensraum NS1. Jeder Knoten in dem Vererbungsbaum repräsentiert einen definierten, benannten Datentyp in der Schemadefinition. Mit dem in der deutschen Anmeldung mit dem amtlichen Aktenzeichen 10351897.5 beschriebenen Verfahren werden für die Namensräume NS1..NS2 jeweils lokale Codes vergeben, die in Figur 3 durch die Zahlen links neben den Knoten spezifiziert sind. Diese sogenannten lokalen Typecodes adressieren alle Typen in einem Namensraum eindeutig. Bei der erfindungsgemäßen Signalisierung eines Datentyps ausgehend von einem Startbasistyp OBT im ersten Namensraum NS1 ist die Menge der adressierbaren Typen im zweiten Namensraum NS2 eine - durch die gestrichelte Umrandung angedeutete - Teilmenge TM aller Typen in dem Namensraum NS2. Dementsprechend werden nur wenige Typecodes verwendet, die in Figur 3 durch die Zahlen rechts neben den Knoten spezifiziert sind und die gegebenenfalls nur eine kürzere binärer Repräsentation erfordern.

In Figur 4 sind vier Namensräume NS1..NS4 dargestellt, zwischen denen auch indirekte Vererbungsbeziehungen existieren;

d.h. Vererbungsbeziehungen, bei denen zwischen einem abgeleiteten Typ und einem Basistyp mindestens ein weiterer Namensraum der Namensräume NS1..NS4 liegt.

- 5 Erfindungsgemäß lässt sich nun ausgehend von dem hervorgehoben dargestellten Startbasistyp OBT im ersten Namensraum NS1 eine - in der Darstellung durch die gestrichelte Umrandung hervorgehobene- Menge TM von adressierbaren Typen durch Betrachtung von Vererbungsbeziehungen aller Namensräume
10 NS1..NS4 bestimmen.

In Figur 5 sind für den vierten Namensraum NS4 die gemäß einer Variante des erfindungsgemäßen Verfahrens strukturierten Vererbungsinformationen BT1..BT3 dargestellt.

15

- Zu erkennen ist dabei, dass die Strukturierung derart herbeigeführt wird, dass für jeden Datentyp, der direkt von einem Datentyp aus einem anderen Namensraum vererbt ist (die jeweilige direkte Vererbung ist durch jeweils einen nichtgestrichelten rückwärts gerichteten Pfeil dargestellt), die Vererbungsinformationen BT1..BT3 gespeichert und/oder von einem
20 ersten zu einem zweiten Gerät übertragen werden, wobei die Vererbungsinformationen BT1..BT3 eines Datentypen gemäß der gezeigten Ausführungsvariante aus einer Kennzeichnung des jeweiligen Namensraums NAMENSRAUM_ID und dem lokalen Typcode der Basistypen LBT, LBT' in den jeweiligen vererbenden Namensraum NS1, NS3 sowie NS4 bestehen.

- Ausgehend von diesen strukturierten Vererbungsinformationen
30 BT1..BT3 wird der Kern des erfindungsgemäßen Verfahrens deutlich, der darin liegt, dass statt des gesamten Vererbungsbaums, der sich aus der Vereinigungsmenge der Vererbungsverhältnisse aller Namensräume ergibt, lediglich die Namensräume von Startbasistypen OBT, der zu adressierenden Typen und die
35 Vererbungsbeziehung BT1..BT3 zwischen demjenigen Namensraum (erbender Namensraum), der die adressierbaren Typen enthält, und demjenigen Namensraum (vererbender Namensraum), der den

jeweiligen Startbasistypen OBT enthält, gespeichert und für die Bestimmung der adressierbaren Datentypen TM verwendet werden.

5 Die Vererbungsinformation (-beziehung) BT1..BT3 identifiziert somit im Grunde Basistypen von Kopftypen eines Namensraums, wobei man unter Kopftypen diejenigen Datentypen versteht, die
 10 direkt auf einem Basistyp aus einem vererbenden Namensraum basieren und wobei als Basistypen hierbei diejenigen Datentypen LBT eines vererbenden Namensraums NS1, NS3 sowie NS4 verstanden werden, die

- a) direkter Basistyp (LBT) eines Kopftypen (HT), oder
- b) direkter Basistyp (LBT') eines Kopftypen (HT') eines
 15 etwaigen weiteren sich in der Vererbungshierarchie befindenden - gemäß dem Ausführungsbeispiel dritten- Namensraums NS3 sind, wobei der Kopftyp HT' direkter oder indirekter Basistyp des Kopftypen HT im Namensraum des abgeleiteten Typen ist.

20 Ausgehend von dieser erfindungsgemäßen Strukturierung werden nun bei dem erfindungsgemäßen Verfahren ausgehend von einem Startbasistyp OBT in einem ersten Namensraum NS1 die adressierbaren Datentypen TM in einem zweiten Namensraum NS2 identifiziert, indem

- a) Kopftypen HT in dem zweiten Namensraum identifiziert werden, zu denen Typen LBT und LBT' aus dem ersten Namensraum als Vererbungsinformation BT1..BT3 abgelegt sind, und
- 30 b) der Typ OBT ein Basistyp der Typen LBT bzw. LBT' ist.

Somit lässt sich lediglich zusammen mit dem Vererbungsbaum des erbenden Namensraums ausgehend von einem als Basistyp festzulegenden Datentyps des vererbenden Namensraums NS1 eine
 35 Menge der adressierbare Typen bestimmen, ohne Kenntnis aller im erbenden Namensraum importierter Namensräume haben zu müssen. Hieraus ergibt sich eine enorme Aufwandsminimierung, die

sich in einer Einsparung von Rechenleistung und Speicherplatzbedarf sowie einer beschleunigten Codierung bzw. Decodierung niederschlägt.

Patentansprüche

1. Verfahren zur Codierung eines strukturierten, insbesondere XML-basierten Dokuments, bei dem eine Vielzahl von Codes mittels eines oder mehrerer Namensräume (NS1..NS4) erzeugt und für mittels Namensräumen (NS1..NS4) definierten Typen vergeben werden
dadurch gekennzeichnet,
dass eine Teilmenge (TM) bestehend aus adressierbaren Typen eines Namensraums (NS2) ausgehend von einem Startbasistypen (OBT) auf Grundlage einer Vererbungsbeziehung (BT1..BT3) zwischen den Namensräumen (NS1..NS4) sowie der Vererbungsbeziehungen in einem Namensraum des Basistypen (OBT) und der Vererbungsbeziehungen in dem Namensraum der Teilmenge (TM) ermittelt wird.
2. Verfahren zur Codierung eines strukturierten, insbesondere XML-basierten, Dokuments nach dem Oberbegriff des Anspruchs 1 oder nach dem Anspruch 1,
dadurch gekennzeichnet,
dass zu jedem Namensraum eine Zuordnung zu weiteren Namensräumen derart gebildet wird, dass zumindest eine Zuordnungsinformation (NAMENSRAUM_ID, TYPECODE) dergestalt erzeugt wird, dass zumindest eine Vererbungsbeziehung (BT1..BT3) zwischen einem erbenden Namensraum (NS2) und vererbenden Namensräumen (NS1, NS3, NS4) beschrieben ist.
3. Verfahren nach Anspruch 2,
dadurch gekennzeichnet,
dass die Zuordnungsinformation des erbenden Namensraums (NS2) aus einer Liste von Codes (TYPECODES) der Basistypen (LBT, LBT') von Kopftypen (HT) des erbenden Namensraums gebildet wird, wobei Basistypen (LBT, LBT') Typen sind, von denen direkt der Kopftyp (HT) entstammt (LBT) oder von denen ein Kopftyp (HT') entstammt, der wiederum Basistyp eines Kopftyps (HT) des erbenden Raums (NS2) ist (LBT').

4. Verfahren nach Anspruch 3,
dadurch gekennzeichnet,
dass die adressierbare Teilmenge(TM) ausgehend von einem
Startbasistyp (OBT) durch Bestimmung der Basistypen (LBT,
5 LBT') des vererbenden Namensraums (NS1) ermittelt wird.
5. Verfahren nach Anspruch 4,
dadurch gekennzeichnet,
dass
- 10 a) ausgehend von dem Startbasistyp (OBT) zur Ermittlung der
Teilmenge (TM) Kopftypen (HT) im erbenden Namensraum (NS2)
ermittelt werden, zu denen Basistypen (LBT, LBT') aus dem
vererbenden Namensraum (NS1) durch die Zuordnungsinforma-
tion (NAMENSRAUM_ID, TYPECODE) identifiziert sind,
- 15 b) der Startbasistyp (OBT) ein Basistyp der Basistypen (LBT,
LBT') des vererbenden Namensraums (NS1) ist.
6. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
- 20 dass die den erbenden Namensräumen (NS2) zugeordnete Zuord-
nungsinformation (NAMENSRAUM_ID, TYPECODE) gemeinsam mit dem
jeweiligen Namensraum (NS2) in einem die Codierung und/oder
Decodierung vornehmenden ersten Gerät gespeichert wird.
- 5 7. Verfahren nach Anspruch 6,
dadurch gekennzeichnet,
dass die den erbenden Namensräumen (NS2) zugeordnete Zuord-
nungsinformation (NAMENSRAUM_ID, TYPECODE) in einem zweiten
Gerät erzeugt und gemeinsam dem jeweiligen Namensraum (NS1
30 NS2) in einem die Codierung und/oder Decodierung vornehmenden
ersten Gerät übermittelt wird.
8. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
- 35 dass für ein Schema und/oder einen Namensraum und/oder für
eine Gruppe von Schemas und/oder Namensräumen jeweils separa-
te, von anderen Schemas und/oder Namensräumen unabhängige Co-

des für die mittels in den Schemas und/oder Namensräumen und/oder in den Gruppen von Schemas und/oder Namensräumen definierten und/oder deklarierten Elemente vergeben werden.

- 5 9. Verfahren nach Anspruch 8, bei dem zur Identifikation des Schemas und/oder des Namensraums und/oder der Gruppe von Schemas und/oder Namensräumen die separaten Codes in entsprechende Adressbereiche unterteilt sind.
- 10 10. Verfahren nach einem der Ansprüche 8 bis 9, bei dem die separaten Codes jeweils einen lokalen Code bezüglich des Schemas und/oder des Namensraums und/oder bezüglich der Gruppe von Schemas und/oder Namensräumen und einen Identifikationscode zur Identifikation des Schemas und/oder des Namensraums und/oder der Gruppe von Schemas und/oder Namensräumen umfassen.
- 15 11. Verfahren nach einem der Ansprüche 8 bis 10, bei dem separate Codes für globale Elemente und/oder SubstitutionGroups und/oder Datentypen erzeugt werden.
- 20 12. Verfahren nach Anspruch 11, bei dem separate Codes für Datentypen TypeCodes derart erzeugt werden, dass innerhalb des Vererbungsbaums eines Namensraums, der zu einem ersten Datentyp in demselben Namensraum benachbarten Datentyp einen Codeabstand zu dem ersten Datentyp hat, welcher der Anzahl der in diesem Namensraum von dem ersten Datentyp abgeleiteten Datentypen entspricht.
- 30 13. Verfahren nach einem der Ansprüche 8 bis 12, bei dem die separaten Codes innerhalb eines gegebenen Namensraums gemäß einem Verfahren vergeben werden, das folgende Schritte umfasst:
 - in einem ersten Schritt werden alle Datentypen eines Namensraums, die von Datentypen anderer Namensräume vererbt
 - 35 worden sind, in der im MPEG-7 Standard definierten Reihenfolge der globalen TypeCodes der jeweiligen Basisdatentypen in

einer Liste sortiert, wobei die Basisdatentypen die Datentypen in anderen Namensräumen sind, von denen die sortierten Datentypen vererbt worden sind;

- in einem zweiten Schritt werden jeweils diejenigen Datentypen eines Namensraums, die von einem bestimmten Basisdatentypen eines bestimmten anderen Namensraums vererbt worden sind, lexikographisch sortiert;
- in einem dritten Schritt werden alle Datentypen eines Namensraums, die nicht von einem Datentypen eines anderen Namensraums vererbt worden sind, entsprechend der im MPEG-7 Standard definierten Reihenfolge in die bestehende Liste von Datentypen einsortiert;
- in einem vierten Schritt werden die separaten Codes in der Reihenfolge der Liste an die Datentypen des Namensraums vergeben.

14. Verfahren zur Decodierung eines strukturierten Dokuments, insbesondere eines XML-basierten Dokuments, wobei das Verfahren derart ausgestaltet ist, dass ein mit einem Verfahren gemäß einem der vorhergehenden Ansprüche codiertes Dokument decodiert wird.

15. Verfahren nach Anspruch 14 zur Decodierung eines gemäß dem Verfahren des Anspruchs 12 codierten Dokuments, bei dem zur Decodierung eines binären TypeCodes aus der Anzahl der abgeleiteten Datentypen die Codelänge der separaten Codes der binären Typecodes bestimmt wird.

16. Verfahren nach Anspruch 14 oder 15 zur Decodierung eines gemäß dem Verfahren des Anspruchs 5 codierten Dokuments, bei dem zur Decodierung eines bestimmten TypeCodes der Subbaum des Vererbungsbaums des Namensraums, in dem sich der bestimmte Typecode befindet, anhand der Codeabstände zwischen benachbarten Datentypen ermittelt wird.

17. Verfahren zur Decodierung eines strukturierten Dokuments, insbesondere eines XML-basierten Dokuments, wobei das Verfah-

ren derart ausgestaltet ist, dass ein mit einem Verfahren gemäß einem der vorhergehenden Ansprüche codiertes Dokument decodiert wird.

- 5 18. Verfahren nach Anspruch 17 zur Decodierung eines gemäß dem Verfahren des Anspruchs 12 codierten Dokuments, bei dem zur Decodierung eines binären TypeCodes aus der Anzahl der abgeleiteten Datentypen die Codelänge der separaten Codes der binären Typecodes bestimmt wird.
- 10 19. Verfahren nach Anspruch 17 oder 18 zur Decodierung eines gemäß dem Verfahren des Anspruchs 12 codierten Dokuments, bei dem zur Decodierung eines bestimmten Typecodes der Subbaum des Vererbungsbaums des Namensraums, in dem sich der bestimmte Typecode befindet, anhand der Codeabstände zwischen benachbarten Datentypen ermittelt wird.
- 15 20. Verfahren nach Anspruch 17 bis 19 zur Decodierung eines gemäß dem Verfahren des Anspruchs 12 codierten Dokuments, bei dem zur Bestimmung der Basistypen (LBT, LBT'), die von einem Startbasistypen (OBT) entstammen, anhand der Codeabstände zwischen benachbarten Datentypen ermittelt wird.
- 20 21. Verfahren nach Anspruch 17 bis 20 zur Decodierung eines gemäß dem Verfahren des Anspruchs 12 codierten Dokuments, bei dem zur Bestimmung der Anzahl an Typen in der Teilmenge TM ausgehend von den Kopftypen (HT) anhand der Codeabstände zwischen benachbarten Kopftypen (HT) ermittelt wird.
- 30 22. Codiervorrichtung, welche derart ausgestaltet ist, dass ein Codierverfahren nach einem der Ansprüche 1 bis 13 durchführbar ist.
- 35 23. Decodiervorrichtung, welche derart ausgestaltet ist, dass ein Decodierverfahren nach einem der Ansprüche 14 bis 21 durchführbar ist.

200403190

24. Codier- und Decodiervorrichtung umfassend eine Codiervorrichtung nach Anspruch 22 und eine Decodiervorrichtung nach Anspruch 23.

Zusammenfassung

Verfahren und Vorrichtung zur Codierung und Decodierung von strukturierten Dokumenten

5

Die Erfindung betrifft ein Verfahren zur Codierung eines strukturierten Dokuments, bei dem werden eine Vielzahl von Codes mittels eines oder mehrerer Schemas und/oder Namensräume erzeugt. Dabei werden für ein Schema und/oder einen Namensraum und/oder für eine Gruppe von Schemas und/oder Namensräumen jeweils separate, von anderen Schemas und/oder Namensräumen unabhängige Codes für die mittels in den Schemas und/oder Namensräumen und/oder in den Gruppen von Schemas und/oder Namensräumen definierten und/oder deklarierten Elemente vergeben. Des Weiteren betrifft die Erfindung ein hierzu entsprechendes Decodierverfahren, eine Codiereinrichtung, eine Decodiereinrichtung sowie Codier-/Decodiereinrichtung.

10

15

Fig. 5

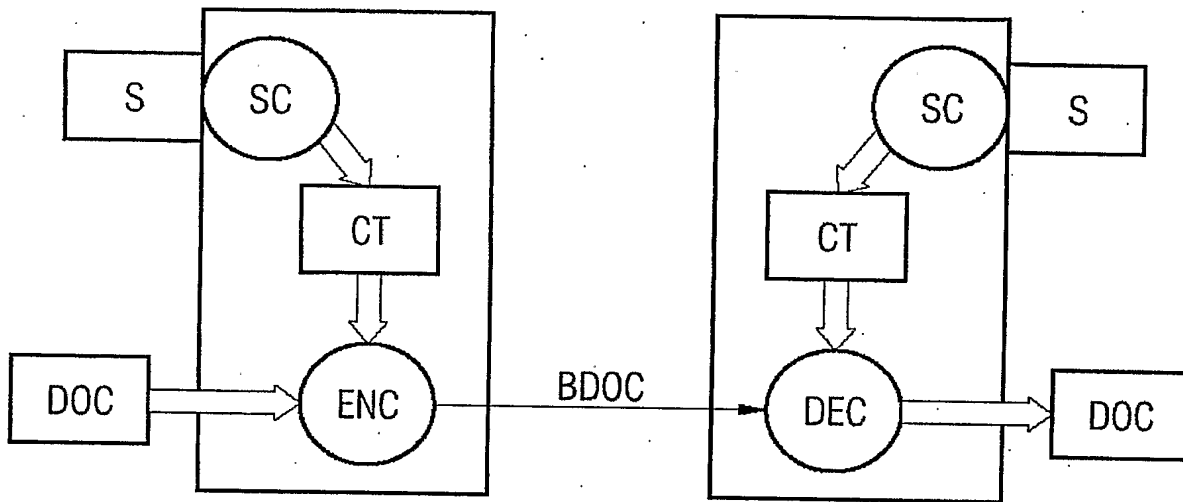


FIG 1

2/5

```

?xml version="1.0" encoding="ISO-8859-1"?
<schema targetNamespace="urn:siemens:SchemaA"
  xmlns:a="urn:mySchema:SchemaA"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  <complexType name="Node1">
    <sequence>
      <element name="name" type="xs:string"/>
      <element name="value" type="xs:integer"/>
    </sequence>
  </complexType>
  <complexType name="myFirstType">
    <complexContent>
      <extension base="a:Node1">
        <sequence>
          <element name="state" type="xs:string"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
  <complexType name="mySecondType">
    <complexContent>
      <extension base="a:myFirstType">
        <sequence>
          <element name="id" type="xs:ID"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
</schema>
<?xml version="1.0" encoding="ISO-8859-1"?
<schema targetNamespace="urn:mySchema:SchemaX"
  xmlns:x="urn:mySchema:SchemaX"
  xmlns:a="urn:mySchema:SchemaA"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  <import namespace="urn:mySchema:SchemaA" SchemaLocation="TreeExample_nsA.xsd"/>
  <complexType name="tType">
    <sequence>
      <element name="name" type="xs:string"/>
      <element name="value" type="xs:integer"/>
    </sequence>
  </complexType>
  <complexType name="iType">
    <complexContent>
      <extension base="a:Node1">
        <sequence>
          <element name="state" type="xs:string"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
  <complexType name="sType">
    <complexContent>
      <extension base="a:Node1">
        <sequence>
          <element name="id" type="xs:ID"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
  <complexType name="nType">
    <complexContent>
      <extension base="x:sType">
        <sequence>
          <element name="price" type="xs:float"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
</schema>

```

FIG 2

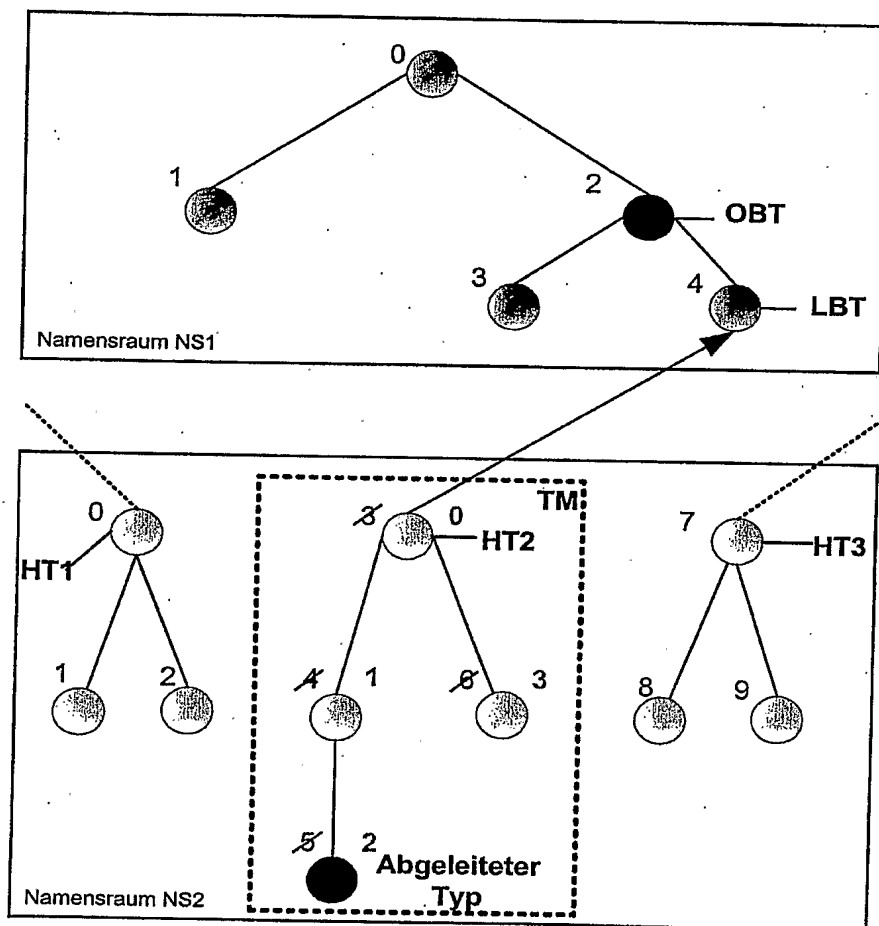


FIG 3

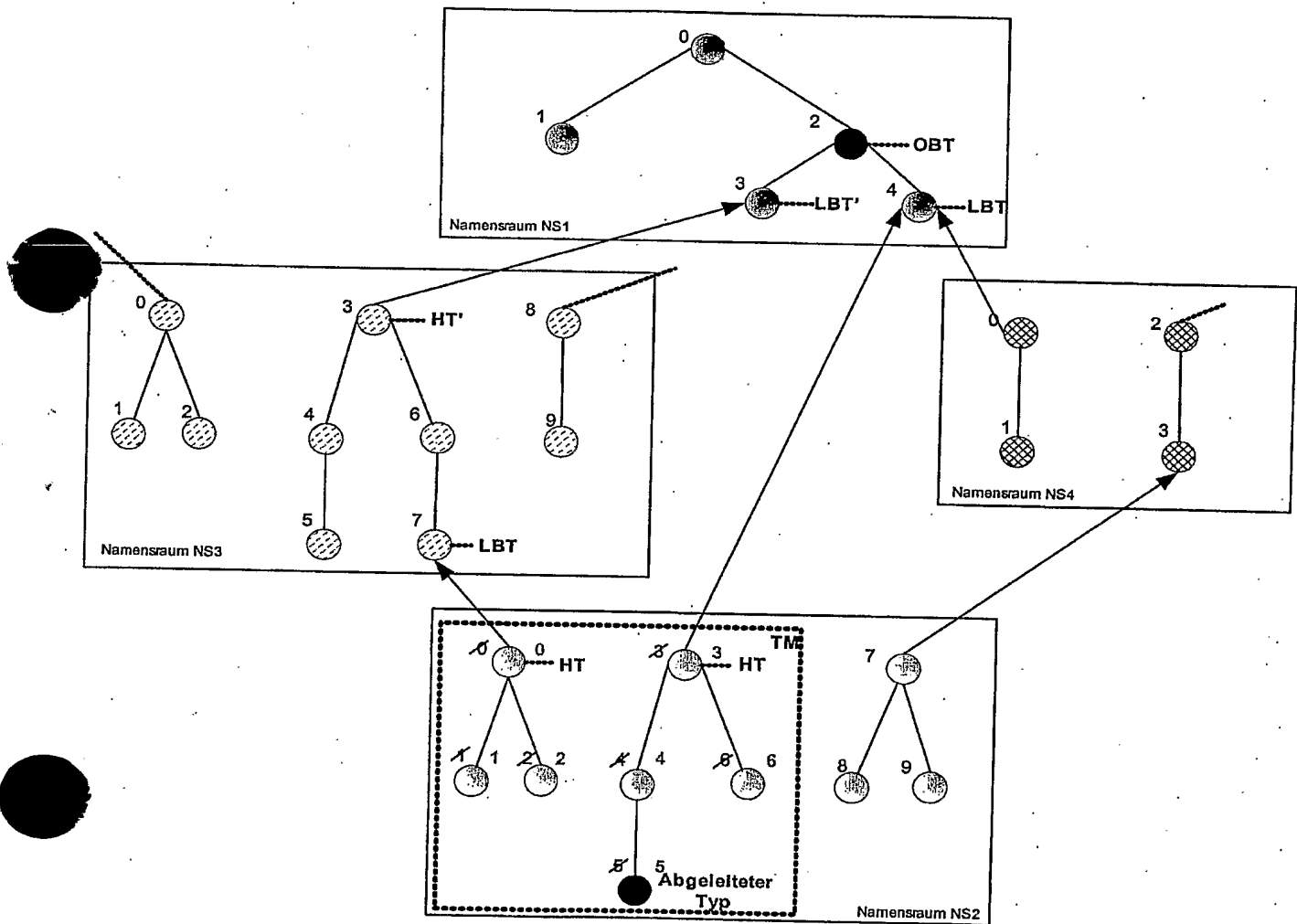


FIG 4

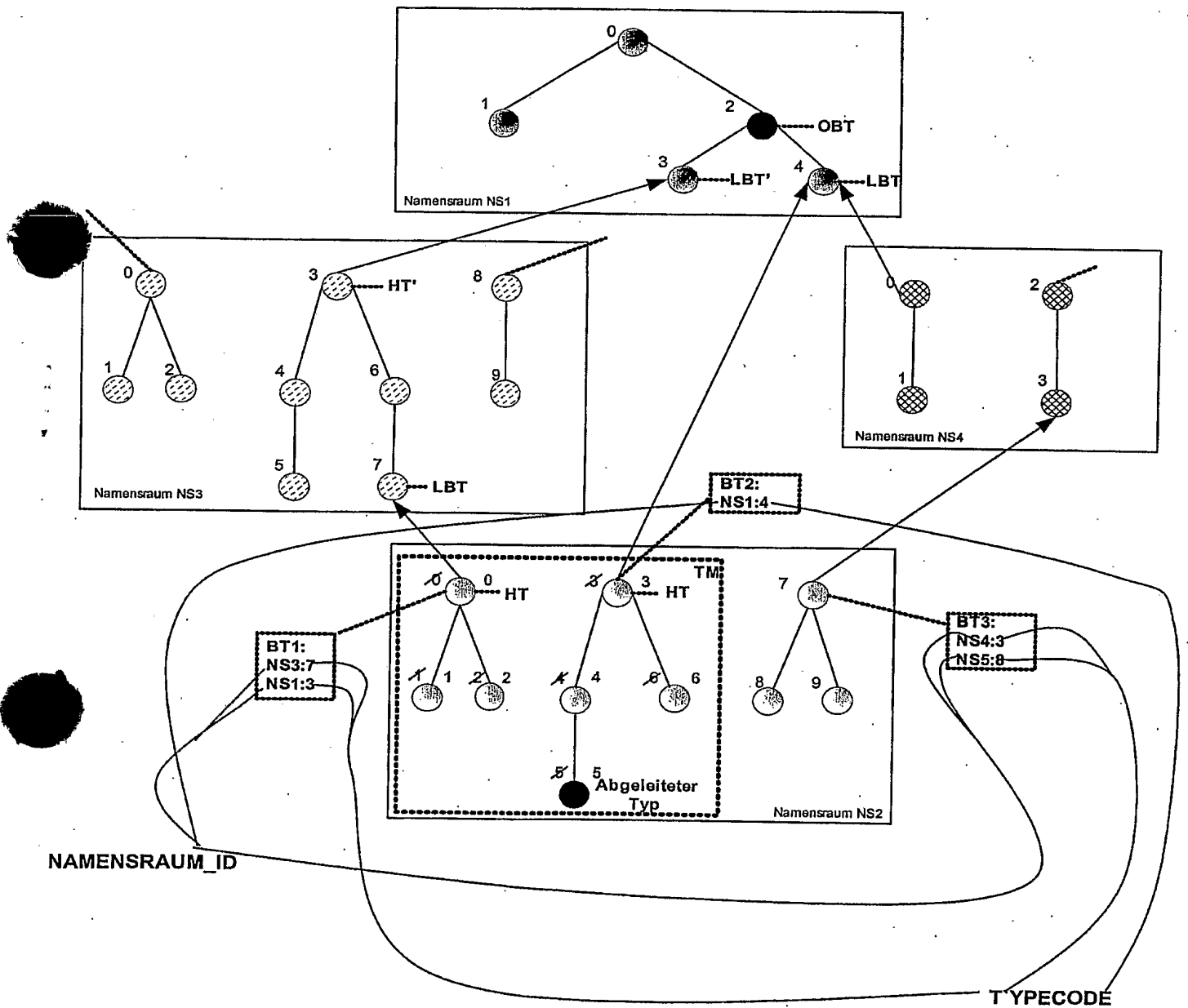


FIG 5